A Novel Method to Gauge Audience Engagement with Televised Election Debates through Instant, Nuanced Feedback Elicitation

Conference or Workshop Item

A Novel Method to Gauge Audience Engagement with Televised Election Debates through Instant, Nuanced Feedback Elicitation

Full Paper

Anna De Liddo
Knowledge Media Institute
The Open University
Milton Keynes, UK MK7 6AA
anna.deliddo@open.ac.uk

Brian Plüss
Knowledge Media Institute
The Open University
Milton Keynes, UK MK7 6AA
brian.pluss@open.ac.uk

Paul Wilson
School of Design
University of Leeds
Leeds, UK LS2 9JT
texpw@leeds.ac.uk

ABSTRACT

Despite a steep increase in the use of the Internet and handheld computing devices for media consumption, television is still of critical importance for democratic citizenship. Television continues to be the leading source of political information and its relevance has been recognised at policy level. In addition, television keeps evolving technologically and in how it is experienced by viewers. Nonetheless, the ways researchers have measured audience engagement with televised political events in real-time is often limited to small samples of viewers and is based upon a narrow range of responses.

In this paper we look at the audience of televised election debates, and propose a new method to gauge the richness and variety of citizens' real-time responses at scale by capturing nuanced, non-intrusive, simple and measurable audience feedback. We report on a paper prototype experiment, in which we used a set of flashcards to test the method in an actual televised election debate scenario. We demonstrate how the method can improve our understanding of viewer responses to the debaters' performances, to the contents in their arguments, and to the debate as media event. We conclude with design guidelines to implement the method on a mass scale in order to measure audience engagement with televised election debates in distributed contexts through audience feedback web and mobile applications.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in interaction design**; *Social engineering (social sciences)*;

KEYWORDS

Televised election debates, audience research, audience feedback, collective intelligence, multimedia annotation, citizen engagement

ACM Reference format:
Anna De Liddo, Brian Plüss, and Paul Wilson. 2017. A Novel Method to Gauge Audience Engagement with Televised Election Debates through Instant, Nuanced Feedback Elicitation. In *Proceedings of C&T '17, Troyes, France, June 26-30, 2017,* 10 pages.
https://doi.org/http://dx.doi.org/10.1145/3083671.3083673

C&T '17, June 26-30, 2017, Troyes, France

ACM ISBN 978-1-4503-4854-6/17/06…$15.00
https://doi.org/http://dx.doi.org/10.1145/3083671.3083673

1 INTRODUCTION

The widespread emergence of ubiquitous personal computing devices and social media are transforming the way citizens experience complex public events. For instance, people engagement with televised political debates is progressively shifting form 'passive' viewing of a television programme to 'active' participation in a wider debate around a televised event [9].

Social media technologies are key players in this change, because they multiply the information channels available to the public around televised political events. These channels are bidirectional: they not only bring information to the citizens but also allow them to play an active role and contribute themselves to the flow. But are these new 'participation experiences' really informative? To what extent do they improve the confidence of citizens on the issues discussed? And most importantly, do the voices in social media truly capture the vast richness and complexity of citizens' reactions to televised political events? What could we learn about the audience of election debates and about the debates as media events if we had better analytical tools to scrutinise audience responses?

This paper reports on a research strand devoted to the study of the audience of televised election debates, and of the ways in which viewers can be encouraged to engage proactively with the debate-viewing experience. We propose a new method to capture the richness and variety of instant citizen reactions to televised election debates, which are difficult to extract with existing techniques based on one-dimensional instant polling. Building on collective intelligence approaches [8] we developed and tested a novel method to harness instant audience feedback in terms of how viewers engage aesthetically, emotionally, intellectually, and critically with televised debates. We then used this feedback to understand the needs of citizens, and to assess election debates as media events.

2 BACKGROUND AND MOTIVATION

Social media are changing the way in which people watch television. While audiences have always been active in their consumption of news and political debates – commenting and talking to others as they watch –, social media networks like Twitter and Facebook enable viewers to join up their conversations and force media presenters and producers to acknowledge this constant flow of live feedback. In the context of televised election debates, this

allows people to share critical responses to implausible claims and inauthentic performances, and to form collective judgements before 'experts' have a chance to influence them. The added sensory streams afforded by social media enlarge the audience experience, but do they help viewers and voters to arrive at informed decisions? Do they help voters to arrive at better decisions? Do they capture the richness of citizens' diverse reactions to complex events like election debates?

Research suggests that existing social media shape and increase the civic engagement compared to offline political engagement [23], especially among young people [24]. These studies mostly refer to the general engagement with politics but not necessarily contribute to gauging a community's reactions to live media events.

Lemert [13] explores the relevance and impact of TV presidential debates in the US from the perspective that they inform a wider spectrum of voters. He presents an analysis of pre-debate and post-debate surveys from the 1988 US campaign and shows that exposure to presidential debates predicted some types of knowledge gain. This finding supports the importance of televised election debates as key democratic moments in which citizens have a chance to engage with and inform their political choices. At present, however, interactions with televised election debates are rather limited. People can engage live with election debates in two main ways: either via polling systems, which are constrained to very small samples of the population; or via synchronous social media such as twitter, at a larger scale.

As said, research shows that social media and the increasing usage of mobile applications can help bridge the democratic divide and improve citizen participation [17]. Nonetheless, polling companies are still considered the most reliable sources of information on audience engagements with televised election debates. These polling experiences are, however, limited in many ways. In the UK, for instance, the House of Lords Select Committee on Communications stated that *the worm* – 'a squiggly line that often accompanies televised election debates and is supposed to represent the views of undecided voters, moving up when a candidate says something which the voters endorse, and down when a candidate says something which they don't like' – 'might distort the viewer's perception of the debate' [12, paras. 163–164]. One reason for this is that the sample of undecided voters used to produce the worm is far too small to have any scientific plausibility: e.g. the broadcaster ITV involved only 20 people in 2010, and the BBC involved 12 [12, para. 165]. A second reason is that the worm simply asks people to indicate which performances they find most convincing, without any reference to why that might be so. These methodological shortcomings, coupled with evidence that the worm may prevent people from making independent judgements when superimposed on live broadcast [6] raise serious concerns.

Analyses of Twitter are rather more sophisticated, using sentiment analysis and other techniques to map the changing mood as expressed in tweets published during debates [1, 18, 19]. But here again, little is learned about which aspects of the debaters' performances trigger certain responses or why they do so. Researchers have raise concerns on the quality of social media contributions to the political debate, and on the soundness of the inferences that can be made from analyses of social media data [21]. If instant audience feedback is to be a new fact of political life, we need better tools for harnessing and interpreting what viewers and voters are thinking.

Research on harnessing communities' ideas, work and actions falls under the areas of crowdsourcing and collective intelligence. Collective intelligence (CI) [14, 16] is an umbrella term used to express the augmented functions that can be enabled and emerge by the co-existence of many people in the same environment – in virtual, real-life or blended settings. We look at the media debate as a blended interaction environment, and seek ways of harnessing collective audience feedback to political debates so as to supports the audience's collective reflection, critical thinking and deep understanding of the debate. Specifically, we build on contested collective intelligence (CCI) [8]. These are discourse-based CI approaches in which collective intelligence emerges from new forms of structured debate and deliberation.

As opposed to crowdsourcing and other CI traditions, which build on the aggregation of unaware traces of users actions, CCI looks at harnessing higher-level thinking traces and aims at capturing people's interpretations to support deeper reflection and understanding. CCI approaches are most appropriate to harness the collective intelligence of communities in complex societal processes, such as political election debates, in which there is more than one world view, there are no solutions that fit all, and dialogue and argumentation are needed to better explore the problem setting and reach collective decisions. We take a Contested Collective Intelligence approach in that we aim to elicit higher-level viewer traces, which consist of discourse elements such as claims, ideas, issues, statements and questions.

Collective Intelligence approaches in general, and CCI approaches in particular, aim to improve a community's collective capacity to make sense of complex problems by enabling the following process: harnessing community traces (community sensing), processing and analysing these traces in a way that provides new knowledge and insights (advanced analytics), present these insights back to the community to improve community reflective capability and inform community action (visualisations and new knowledge presentation, the closing up of the CI cycle – Figure 1). De Liddo et al. [7] discuss the role of analytics and visualisations in CI processes.

In this paper we present a method that supports the first two steps of a collective intelligence process: we harness the audience higher-level thinking and understanding of a televised election debate, in a way that enables advanced analytics and insights on the audience's understanding and experience. More specifically, we propose a participatory method to actively involve the audience in a new viewing experience of election debates. The main mechanism of audience engagement we propose consists of enabling instant, nuanced audience reactions to the live broadcast or post-hoc video replay of a televised election debate. Enabling such feedback has four main objectives:

- promoting active engagement by allowing the audience to react to the televised debates in new non-intrusive, yet expressive, and timely manner;
- harnessing viewers' reactions as collective intelligence that can be analysed both in terms of the immediate viewer experience and longer-term shifts in political preferences;

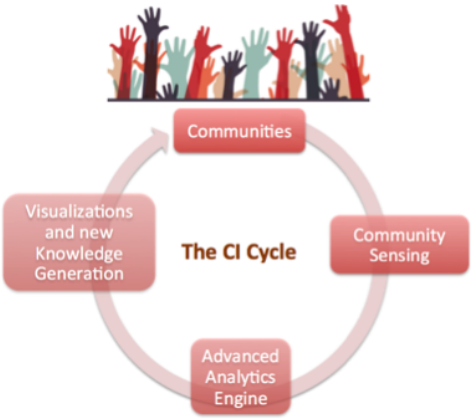

Figure 1: A collective intelligence cycle can be seen as composed of three main steps: community sensing, data aggregation and analytics, visualisations and new knowledge generation to feed back into the community.

- understanding the complex and nuanced nature of collective and individual responses to the debates;
- providing new ways to assess the debates as both 'media events' and 'democratic opportunities', and developing ways of making future televised election debates more cognitively, affectively, critically and aesthetically appealing to voters.

3 INSTANT, NUANCED AUDIENCE FEEDBACK METHOD: CONCEPT AND DESIGN

Based on the objectives laid out above, and following a CCI approach, the audience feedback we propose to capture has the following six main attributes:

- it is provided in form of discourse elements, such as statements, questions, claims, opinions, issues, and arguments;
- it is instant in that is indexed to specific moments in time in the televised debate video;
- it is nuanced in that it captures the rich variety of meaning in viewers' reactions;
- it is non-intrusive to the viewing experience;
- it is easy to understand and use, because it builds on natural language and lay people jargon;
- it is specific to the aspects of political communication we want to measure.

We refer to this concept as 'instant, nuanced feedback'. This is to capture the lightness of its affordance, the coupling with a specific instantaneous fragments of a multimedia event, and the sophistication of the details it captures, which enables gathering a rich picture of audience reactions to televised debates.

To test the concept we designed an experiment in which possible audience reactions were represented by a set of flashcards. The method is inspired by the Leitner system [5, 11], in which the flashcards consist of paper cards with textual information, often used in learning contexts for memory training. In this case, we used paper flashcards to help viewers to reflect on their reactions to the televised event by using the textual prompts presented in each card.

Table 1: Meaning and textual content of the set of flashcard for feedback elicitation

Category	Meaning	Textual Content
Information Need	Fact checking need	*Is this true?*
	Information Need	*Where can I find more info on this?*
	Personal engagement need	*How does this affect me?*
	Civic engagement need	*Why should I care?*
	Trust Need	*Can I trust him?*
	Argument mapping need	*What are the pros and cons of this policy?*
Trust	Distrust the speaker	*He is vague and avoiding the question!*
	Distrust the claim	*This is a wrong statement!*
	Emotional distrust	*I do not believe this!*
	Emotional trust	*I believe in this!*
	Trust on the claim	*Correct!*
	Trust on the speaker	*His response is confident and precise!*
Emotion	Happy	*I love it!*
	Pleased	*This is better than I expected*
	Unenthusiastic	*I'm losing interest*
	Bored	*I would leave the room now if I could*
	Disappointed	*This is so sad!*
	Angry	*This is unnerving!*

The cards are organised in three dimensions: information needs, trust, and emotion. Information need cards are aimed at providing insights on a viewer's gaps in knowledge and at making explicit the points at which the information conveyed by the debaters needs to be complemented. Trust cards are aimed at providing insights on the main motivators for a viewer's trust or distrust on the speaker, on the event, or on their own beliefs. Emotion cards are aimed at providing insights on a viewer's emotive reactions to the debate and can be used as proxies to assess audience engagement with the speakers, the topics under discussion, etc. In order to capture nuanced reactions to each of these dimensions we designed six flashcards for each category. Trust and emotion cards were designed to capture a nuanced polarisation of positive and negative reactions. Information need cards, on the other hand, were not semantically polarised and aimed at capturing key questions that the viewers may have had in relation to different elements of the televised debate. The meanings, and the textual contents of the flashcards we used to capture these meanings, are shown on Table 1.

3.1 Design of the Flashcards

Based on the meanings and textual contents we designed the 18 coloured flashcards shown in Figure 2.

Design Methodology. We followed an iterative process, beginning with the definition of a distinct set of visual 'problems' tied to the distinct visible language needs of participants. Questions such as: What do the cards need to do? and How must the visible language work to help address these needs? allowed for a mapping of possible visual approaches together with key considerations within the designs (format, colour range, etc.). These considerations were explored through idea generation and the scoping of potential visual approaches, with the production of a range of typographic styles and visual systems to formalise the typographic potential for the cards. A refinement of one typographic direction led to the executed approach, deemed most appropriate for the participants.

Figure 2: Deck of flashcard for feedback elicitation: information need (blue), trust (yellow) and emotion (red)

Format. The format of the cards themselves centred around ideas of visual and ergonomical necessity, needing to be easily managed and manageable as participants respond to the debate. Given this functionality, the cards would need a high degree of visual immediacy – requiring both speedy recognition and notable difference to ensure their effective sorting by the participants and, as such, were conceptualised as a kind of hand-tool.

Typography. The typographic form for the cards was designed with emphasis on:

Size. Text size was not fixed but, instead, the approach looked to employ a dynamism of scale across the card and with each typographic unit (word or words) with visual emphasis attributed to key words within each phrase.

Composition. Typographic layout and the visual arrangement of words looked to be dynamic, resulting in each card having a visually distinct 'look'.

Colour. Two factors were key to use and selection of colour: to ease the experience of reading, and to ensure speedy recognition. The colours had strong visual contrast to ensure that they were usefully distinct and to make categorisations easy to recognise. A visual shift in each colour's tone across the card deck was utilised to distinguish individual cards.

Typeface. The use of Berthold Akzidenz-Grotesk aimed to ensure some neutrality within the letterforms themselves, as participants were to be more focused on each card's content.

Symbol. A background mark or visual indicator of each category was used to help participants distinguish between cards. Either question mark or exclamation mark (upon a background colour) would help frame the response.

Softness. In design terms relatively under or un-designed; early versions of the cards made more use of a variety of detailed typefaces of more complex layout and graphic devices.

4 USER STUDY: THE SECOND CLEGG-FARAGE EUROPE DEBATE

Our main hypothesis is that the instant, nuanced audience feedback method is at the same time intuitive and able to provide advanced analytics on the audience and the debate. Specifically, we hypothesized that the use of the audience feedback cards would be easy for untrained users and that the method would allow us to provide a rich overview of the audience understanding and reactions to the debate in form of advanced analytics. We organised a user study to test these hypotheses and gather design guidelines for an online instant audience feedback tool based on the proposed method.

We tested the proposed audience feedback method in the second of two debates between Liberal Democrats Leader and UK Deputy Prime Minister Nick Clegg and UK Independence Party Leader and Member of the European Parliament Nigel Farage on whether the UK should be in or out of the EU[1].

We recruited 15 students from the University of Leeds[2] who watched the debate live in the same room. The participants were each given the pack of 18 flashcards and asked to raise any card in the air if it expressed their thoughts or feelings at any point during the debate. Participants were encouraged to raise any of the cards as often as they wanted. The experiment was video recorded, allowing us to analyse responses at both an individual and group level. The video was then analysed to timestamp and annotate each occasion on which a card was raised.

Participants raised over 1470 flashcards during the one-hour debate. From the annotation process, we draw qualitative insights on the participants' use of the flashcards. We analysed the responses of the entire group, as an illustration of collective intelligence: individual patterns of response, which can be correlated with information regarding individuals' socio-demographic and political profiles, and patterns of group response to the performances of individual debaters. Additionally, we carried out a quantitative analysis of the collected data and produced visual analytics to show how often flashcards were displayed, their distribution over time, and their correlation with what was being said in the debate.

4.1 Data Collection

Participants seated in two rows, in the same room, with the debate projected live in front of them on a large screen. We recorded the entire session using two cameras, ensuring that all participants were fully visible at all times and that the flashcards could be seen and distinguished clearly every time they were raised. We combined the two recording in a single synchronised video as in Figure 3.

In addition, we produced a full timed transcript of the debate. We achieved this in two steps: first, we used YouTube to produce an automatic close caption[3] of the audio of the debate and, second, we manually adjusted the timings and corrected any errors introduced by the automatic speech recognition.

4.2 Data Annotation

The annotation was carried out using the Compendium[4] concept mapping tool [4, 20, 22]. Compendium allows for the creation of

[1]Clegg and Farage held two one-hour debates in March and April 2014 before the UK European Election. The second debate aired live on the BBC on 2 April 2014; http://www.bbc.co.uk/news/uk-politics-26854894.

[2]This was in no sense a representative sample of the UK electorate and we make no claims about the representativeness of their responses. Our sole purpose in this experiment was to test the audience feedback method.

[3]https://support.google.com/youtube/answer/3038280

[4]http://compendium.open.ac.uk/

Figure 3: Setup for the flashcard experiment: 15 students from the University of Leeds were given the entire set of cards and sat in the same room as they watched the entire second Clegg v. Farage Europe debate live on 2 April 2014.

maps – dialogue, argument and issue maps – to be made on a video, with nodes and connection linked to specific points in time of the debate [2].

We used a modified version of the movie mapping functionality of Compendium to make digital annotations of every instance in which a card was raised by one of the participants. These instances were linked to the point in time of the debate in which they happened. The setup for the annotation is shown in Figure 4. There was one static node for each flashcard, and one node for each participant which appeared only when the participant raised a card. These appearances were linked with all the cards raised by the participant so far in the video, providing a visual clue of the annotations and of the relevant entries in Compendium's internal connections database. These connections, linking a participant, one or more cards, and a point in time were later exported to standard formats (e.g. comma separated values) for analysis. They were also exported as XML following the YouTube annotation scheme[5], which could be imported on YouTube for dissemination and collaboration. The annotation of the entire debate took around 20 hours to complete, involving a single researcher. It is important to highlight that there was no subjectivity in the annotation as the researcher was just coding what viewer lifted what card when, all objective observations that would be easily automatically obtained in a digital version of the method.

5 QUALITATIVE INSIGHTS

During the annotation of the video, a number of qualitative insights came to light which are worth of note. The most encouraging – and somewhat surprising – was the ease with which participants engaged with the deck of cards. In fact the cards are not simple red/green cards, they consist of 18 rather complex statements, semantically charged, that participants have to use instantly as they watch and processed a complex media event. Our concerns in connection with the complexity associated to the number of options dissipated as participants seemed to quickly and naturally incorporate the selection and raising of the cards as they viewed the debates. What is more, the level of engagement did not decrease as the debate progressed. Rather on the contrary, in many cases it increased and, by the end of the debate, some participants seemed so comfortable with the cards that they used them humorously

[5]https://www.youtube.com/yt/playbook/en-GB/annotations.html

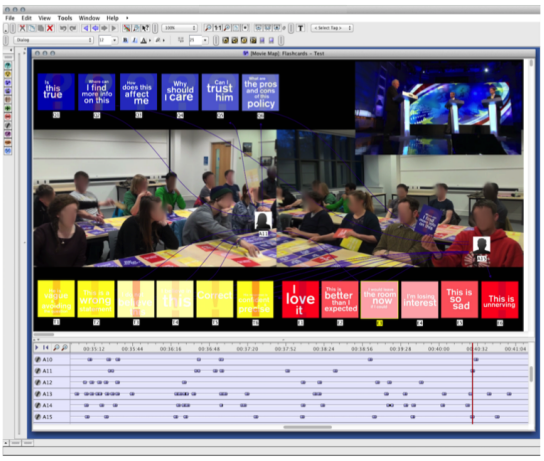

Figure 4: Annotation setup for the flashcard experiment using a special-purpose version of Compendium.

to qualify the facilitator's closing remarks.This is also substantiated by data on the annotations' frequency which remained mostly constant along the debate.

Some of the finer-grained insights we obtained from annotating the data enabled the identification of six affordances of the flashcards:

Housekeeping. Participants proactively reorganised the deck of cards in ways that were more natural to them (see Figure 5), sometimes reordering them to have those they used most often at hand. They also took care of keeping them in order and on sight after each time they raised a card.

Card usage. As the annotation progressed it became evident that the design of the cards was instrumental to the ease with which participants seemed to be using them. After an initial period, the 'shape' of the cards – given by the weight and typography used in the words in a card and the way these are arranged –, combined with the colour allowed for an immediate identification of the card. We hypothesise that this facilitated the rapid selection of suitable cards as the participants reacted to the events in the debate.

Combinations. About 20 minutes into the debate, some participants started using two cards at the same time (see Figure 6). This practice became more frequent until about half the participants engaged in it. On occasion, one of the participants combined three cards in one instant reaction.

Outliers. Quickly into the annotation of the experiment it became evident which of the participants were more enthusiastic about the use of the flashcards and which were not so engaged. In particular, the two outliers in total number of cards raised through the debate were easily identified and later confirmed in the quantitative analysis.

Participant profiles. A remarkable outcome of the annotation was that it made the 'profiles' of each participant rather clear in terms of aspects which would have been otherwise rather difficult to elicit. These included their preference for one of the candidates, political inclinations, whether they favoured social, informational or individual consequences of the points covered in the debate, their tolerance to vagueness and fussy rhetoric, etc.

Figure 5: Participants were allowed to organise their decks of cards as they pleased.

Figure 6: On many occasions, participants raised several cards at once combining them to express more complex responses to the same statements.

Interactions and peer pressure. On many occasions there were interactions between the participants, either bidirectional by conversation or gestures, or by paying attention to the cards raised by others (most often immediate neighbours). These certainly led to an increase in activity and in some cases possibly an alignment or opposition in their reactions.

As mentioned above, most of these findings, although subjective at the time of the annotation, were later confirmed by the quantitative analysis.

6 QUANTITATIVE ANALYSIS

During the debate, participants raised 1472 flashcards. Of these, 701 (48%), were trust-related responses, 575 (39%) were information needs, and 196 (13%) were emotive reactions. Using the timestamps in the debate transcripts and the time-linked annotations of the video of the experiment, we were able to determine to what statements – and therefore to what speaker – the participants were responding to. Of the total number of cards raised, 584 were triggered by Clegg's statements, 765 by Farage's statements and 159 were linked to one of the other participants of the debate – the moderator, members of the audience – or were linked to moments in which it was not clear which one of the debaters was speaking. Of the flashcards raised when Clegg was speaking, 308 (56%) were

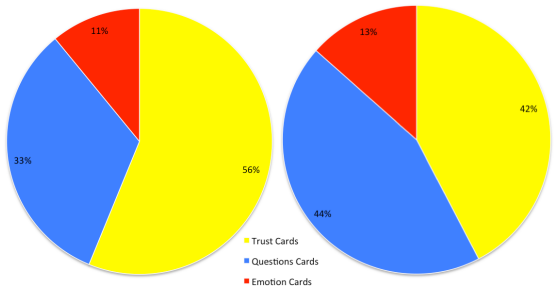

Figure 7: Distribution of reaction triggers per category for Clegg (left) and Farage (right)

Table 2: Distribution of the flashcards lifted by the 15 participants over the one-hour period of the debate.

	Trust	Information Need	Emotion	Total
Clegg	308	180	60	584
Farage	324	338	103	765
Other	69	57	33	159
Total	701	575	196	1472

trust-related, 180 (33%) were information needs, and 103 (11%) were emotive responses. Farage's trust-related triggers were 324 (42%), while 338 (44%) were information needs and 103 (14%) were emotive responses. These values are given on Table 2 and shown graphically on Figure 7.

Initial results of the quantitative analysis show that, overall, Farage provoked more reactions than Clegg – especially information need and emotive responses. When Clegg was speaking he provoked more trust-related reactions than those in the other categories combined. Farage's statements triggered almost as many trust-related responses as information needs (see Figure 7).

Looking in greater detail at the number of times each flashcard was lifted in response to the politicians' statements, the spider diagram at the top of Figure 8 shows that emotive reactions to Farage's contributions were mostly negative, with reactions like *This is so sad!* and *This is unnerving!* accounting for over 70 of the 103 emotive reactions. On the other hand, Nick Clegg seems to have been perceived as trustworthy, as indicated by the spider diagram in the middle of Figure 8. Participants overall *believed in what he said* (85 of the flashcards raised), and found his statements *correct* (65 of the flashcard raised), although in many occasions he was seen as *vague and avoiding the questions* (70 of the flashcard raised). Participants considered Farage's claims were often *not believed to be true* (80 of the flashcard raised) and his statements were often considered *wrong* (also 80 of the flashcard raised). Looking at the information need flashcards on the spider diagram at the bottom of Figure 8, we can see that Clegg's area (light blue contour) is fully contained in Farage's (dark blue lines). The audience questioned almost three time more often the credibility of Farage's statements compared to Clegg's – 160 *Is this true?* flashcards were raised in connection with Farage's claims, compared to only 60 raised in reaction to Clegg's. It is also of interest to notice that, overall, participants almost never wondered how the topics under discussion related

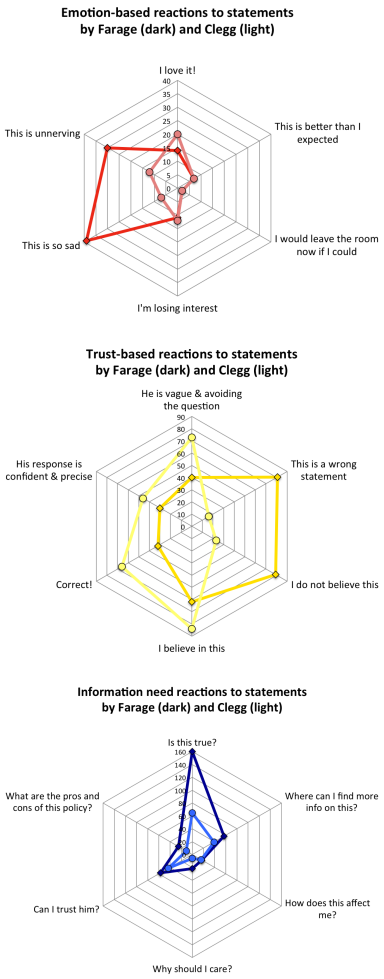

Figure 8: Combined spider diagram of the reactions in each category for Clegg (light colours) and Farage (dark colours)

to their lives or to their civic roles: the flashcards *How does this affect me?* and *Why should I care?* were rarely used. The need for knowing about *the pros and cons of the policy under discussion* was also almost absent. We speculate this could be related to the themes of the debate, which was not focused on policy but on whether the UK would be better in or out of the European Union, leaving little room for the reactions captured by this flashcard.

Focusing on trust-related and emotive responses, we analysed the polarity of the audience responses to each debater's contributions. The trust-related flashcards *He is vague and avoiding the question!*, *This is a wrong statement!*, and *I do not believe this!* were considered negative, while *I believe in this!*, *Correct!*, and *His response is confident and precise!* were considered positive. The emotive responses *I would leave the room now if I could*, *This is so sad!*, and *This is unnerving!* were considered negative; while *I love it!*, *This is better than I expected* and *I'm losing interest* were considered positive[6].

[6]Although *I'm losing interest* expresses a negative change of state, we focused on the implicit positive assertion that the viewer is currently interested - although this interest is declining.

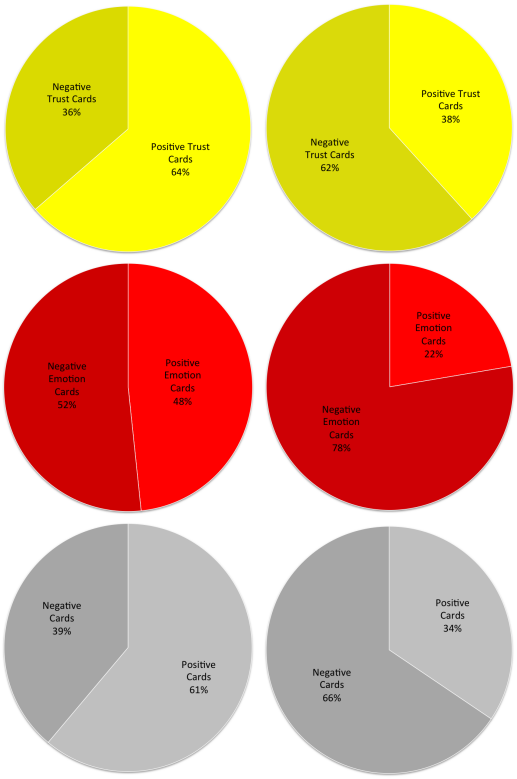

Figure 9: Polarity of emotion and trust reaction triggers for Clegg (left) and Farage (right)

The results are shown on Figure 9 for each one of the speakers, separated into trust-related responses (yellow), emotive reactions (red) and both combined (grey). It is clear from the charts that the reactions to Clegg's statements were significantly more positive than those to Farage's. This came as a surprise, given the consensus in the media immediately after the debate that, according to polls, Clegg had 'lost' the debate by a wide margin[7].

Finally, a timeline representation of all the flashcards raised in the debate was produced as a means to spot critical moments (see Figure 10). By looking for the points along the duration of the debate with higher concentration of flashcard-raising events – visualised as regions with greater colour intensity –, we can identify the fragments of the debate that provoked the stronger reactions in our audience. For instance, between minutes 7 and 8 a spike of blue lines – information need flashcards – is noticeable. Upon closer inspection of the data (see Figure 11), the majority of the flashcards that were raised read *Is this true?* or *Where can I find more info on this?*. Viewers were reacting to the following claim by Farage: 'unless we get reform then the time has come to leave the EU' by wondering whether the statement was true or requesting for sources of further information on the matter. A peak of red and yellow flashcards – emotive and trust-related reactions, respectively – is noticeable between minutes 28 and 29. This coincides with a

[7]The Guardian, for instance, reported that 69% of those polled chose Farage as the winner; http://www.theguardian.com/politics/2014/apr/02/nigel-farage-triumphs-over-nick-clegg-second-debate.

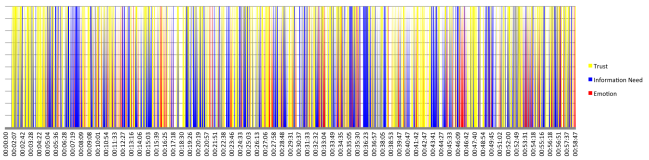

Figure 10: Timeline representation of audience reactions over the duration of the debate

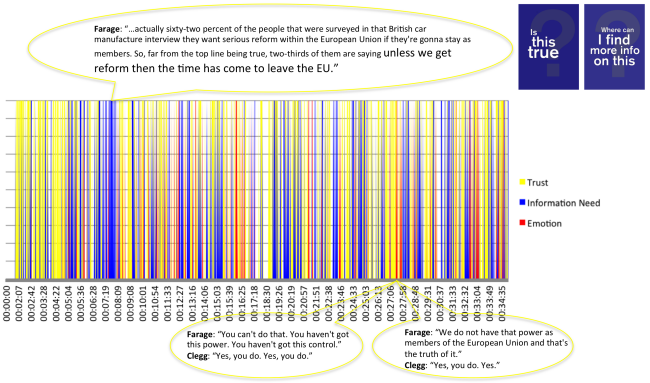

Figure 11: Linking audience feedback to debate moves

period in which Farage and Clegg contradicted basic claims made by one another, suggesting viewers became unsure about whom to trust and irritated by the position in which they were being put.

7 WHAT CAN WE LEARN FROM INSTANT, NUANCED AUDIENCE FEEDBACK?

Initial results from our quantitative analysis showed that, overall, Farage provoked stronger emotive reactions than Clegg and that these were mainly negative. When Clegg was speaking he triggered more trust-related reactions and these were mainly positive. Overall, the participants in our experiment *believed* what Clegg said, found his statements to be valid, even though many of them considered him to be *vague and avoiding questions* at times. Participants expressed doubt about claims made by Farage and his statements were often considered to be *untrue*. From the spider diagram in Figure 8, we saw that the audience questioned the credibility of Farage's statements almost three time more often than they questioned the credibility of Clegg's statements. The analysis of the polarity of trust and emotive audience reactions showed that these were significantly more positive when Clegg was speaking than when Farage was.

Seeking to go beyond questions of affect (how viewers felt about each of the debaters) and trust (how credible debaters' claims seemed to viewers), we were able to elicit what viewers believed they needed to know in order to evaluate the debate effectively. Generally speaking, cards asking for more information about a policy were rarely raised. But as shown by the timeline in Figure 10, we can see that there were several critical moments in which the debate viewers felt confused and wanted more information to help them make sense of what was going on.

The method allows for a better and finer grained understanding of both the political debate and people's reactions to the viewing experience. The narrative reported above represents not only an in-depth understanding of the political debate but also provides quantitative evidence for this interpretation. The method allows for the production of visual analytics to show how often flashcards were displayed, their distribution over time, and their correlation with what was being said in the debate. The responses of the entire audience can be also analysed as an illustration of collective intelligence, plotting individual patterns of response, which can be correlated with information regarding individuals' socio-demographic and political profiles, and patterns of group response to the performances of individual debaters. As an example, Figure 12 shows the flashcard-raising profile of each participant and the group mean (in orange). In addition to providing an insight on the behaviour of the audience as a collective, this representation allows for the detection of outliers and other interesting cases. Moreover, Figure 13, which derives from the previous chart, could be presented to individual viewers as a way of encouraging reflection on their responses with respect to those around them or to the entire group. These are just a few examples of the analytical and interpretation power of the method proposed in this paper.

8 GUIDELINES FOR THE DESIGN OF INSTANT, NUANCED AUDIENCE FEEDBACK WEB AND MOBILE APPLICATION

The use of paper flashcards to harness instant, nuanced audience feedback showed promising results in terms of active engagement and appreciation from the participants of the experiment. Still, paper cards have obvious limitations in terms of the cost of data collection, annotation and analysis, which make a mass scale implementation hard – if not impossible. To overcome this and allow gathering instant feedback from larger audiences, we are designing and developing audience feedback web and mobile applications, following the same principles described in Section 3.

As a first step toward the design of these applications, and as a direct result of the experiment, we translated the insights form the qualitative and quantitative analyses discussed above into a set of guidelines that work as high-level system requirements for harnessing audience feedback at scale and in distributed environments:

(1) the design principles used in the flashcards must be preserved in the digital version;
(2) it must be possible for users to arrange the flashcards as they consider most convenient – before and during the event for which feedback is being elicited;
(3) the system must provide visual feedback when a user has 'raised' a card;
(4) the moment in which a user 'raises' a card with respect to the period of the media event must be recorded by the system without the user's intervention or awareness;
(5) it must be possible for users to indicate the duration of their responses, e.g., by leaving a card 'raised' for an arbitrary period of time before 'putting it down';
(6) users must be allowed to select more than one card at the same time to express more complex responses;

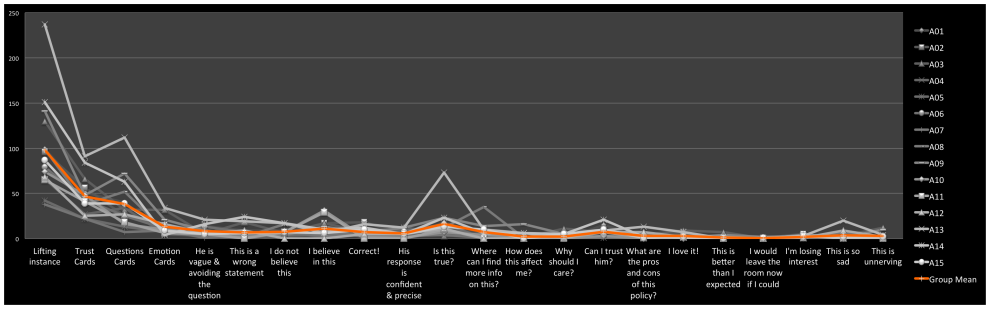

Figure 12: Performance of audience members versus the group mean for the detection of outliers and other interesting cases

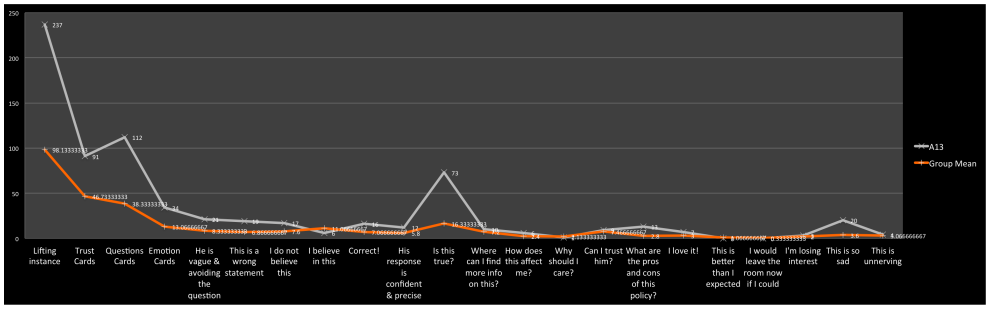

Figure 13: Performance of an individual with respect to the group mean

(7) in order to mimic the effect of other members of the audience in encouraging engagement with the debate and with the feedback system, users should be exposed to a stream of 'peer activity' showing the responses of other viewers.

Web and mobile applications are being designed and implemented following these guidelines, in an attempt to replicate the successful design choices from the paper flashcard experiment. Figure 14 shows a very simple web application prototype. The meanings, textual contents and graphic design – colours, layout, and typography – are currently being revised and re-designed based on the results of the experiment in order to capture new dimensions of the political communication and democratic process.

We plan to replicate the instant, nuanced feedback elicitation in an A-B testing experiment in which we compare the implementation of the method in a virtual distributed setting (individual viewers using the web and mobile applications while watching a debate on their own), and in a group face-to-face setting (a group of viewers using individual digital versions of the flashcards while watching a debate in the same room). The study is aimed at comparing the strengths and weaknesses of the application of the method in face-to-face vs. virtual contexts. The main goal of this research effort is to provide guidelines on how to use the instant, nuanced feedback method to harness audience reactions to televised election debates in different contexts, and with audiences of different sizes.

9 CONCLUSION AND FUTURE WORK

A vast literature already argues that the use of social media enhances and expands citizen engagement in politics [17, 23, 24]. This motivates our research on developing and applying new and

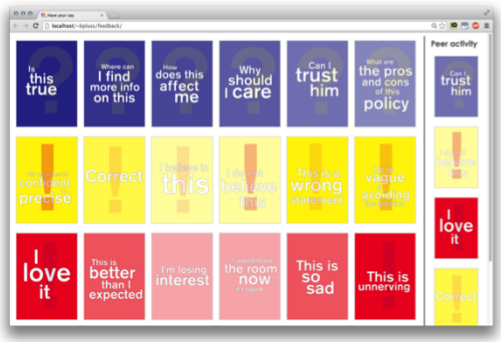

Figure 14: A web application prototype of the flashcards

common media technologies to improve citizens' participation in different political events, such as televised election debates.

Literature on real-time analysis of audience behaviours during live events (e.g. elections, sports, disasters), especially applied to sentiment analysis of Twitter activity, shows that rather sophisticated selection and tracking of users are already possible [15]. These methods are relevant in that they also allow the segregation of viewers' responses and enable tracking users independently. Nonetheless, most sentiment analysis of Twitter streams present outcomes that are still reduced to targeted sentiments and win-lose predictions. Additionally, most qualitative studies of Twitter streams presented in the literature show that qualitative discourse analysis has a great explanatory power on audience reactions (see Brooker et al. [3] for a very interesting analysis of tweets during

politically-charged TV broadcasts). Still, qualitative analyses of popular tweets raise obvious questions of scalability, data sampling and generalization of results. The instant, nuanced feedback method we propose aims to provide similarly powerful insights on the audience while preserving the accountability of the results.

We are currently undertaking research to compare the wide-ranging, often nuanced audience reactions captured by our method with the analysis of live reactions on social media produced by techniques such as sentiment analysis. For instance, Diakopoulos and Shamma [10] presented a study based on the first 2008 US presidential debate in which journalists and political insiders used specific visuals and metrics to understand the evolution of sentiment in reaction to a debate video. The method proposed by the authors is similar to the instant audience feedback method in terms of predictive power but incorporates viewers via established social media. We plan to test whether social media analysis and our method identify similar critical events within the same televised election debate, and compare the types of insight that can be inferred by using either method to assess audience engagement, understanding, and appreciation of the debate. This research will contribute to advance our understanding of the potentials and limitations of using novel as well as and common social media technologies for studying citizen engagement with such media events.

As we said, the experiment that we have described does not scale up. The use of paper flashcards in a face-to-face setting can only capture a small number and range of reactions to a televised debate and form a small number of participants. The number and type of people in our sample cannot be regarded as representative of a wider population. This is why it is important to repeat that we draw no political conclusions from what our participants told us: our sole aim was to test a methodological concept. During the UK 2015 ITV General Election debate, we developed a web application to test the method by gathering responses from a larger, more representative sample of debate viewers. We are currently developing a statistical method to analyse how viewers' demand for specific democratic capabilities and entitlements were triggered at specific moments in the debate in relation to particular topics and themes. We shall produce analyses at both a macro (aggregate) level and a micro (individual) level. Our aim will then be to refine the method and make it available for use in other contexts via a mobile application.

Finally, in terms of contributing to research on Collective Intelligence we have to highlight that the method we propose only enables the first two steps of a Collective Intelligence cycle (community sensing and advanced analytics). Future research needs to address the problem of linking our audience feedback harnessing method with a fully-fledged collective intelligence technology. A collective intelligence technology built on the instant, nuanced audience feedback method would have to automatically enable and inform the entire CI cycle, including the automatic development of the analytics from the audience feedback, and the study of how these analytics should be presented back to the community to trigger reflection and to support knowledge co-creation.

ACKNOWLEDGMENTS

We would like to thank our colleagues in the Election Debate Visualisation project, Prof Stephen Coleman, Prof Simon Buckingham Shum and Dr Giles Moss for constructive feedback and discussions that greatly improved the method. This research was funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant EP/L003112/1.

REFERENCES

[1] N. Anstead and B. O'Loughlin. The emerging viewertariat and bbc question time: Television debate and real-time commenting online. *The International Journal of Press/Politics*, 2011.

[2] H. Bailey, M. Bachler, S. Buckingham Shum, A. Le Blanc, S. Popat, A. Rowley, and M. Turner. Dancing on the grid: using e-science tools to extend choreographic research. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1898):2793–2806, 2009.

[3] P. Brooker, J. Vines, S. Sutton, J. Barnett, T. Feltwell, and S. Lawson. Debating poverty porn on twitter: Social media as a place for everyday socio-political talk. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3177–3186. ACM, 2015.

[4] J. Conklin, A. Selvin, S. Buckingham Shum, and M. Sierhuis. Facilitated hypertext for collective sensemaking: 15 years on from gibis. In *Proceedings of the 12th ACM conference on Hypertext and Hypermedia*, pages 123–124. ACM, 2001.

[5] H. Davatgar and Z. Ghorbanzadeh. The effect of leitner's learning box on the improvement of vocabulary teaching and learning (case study: First year students in parsabad moghan branch, islamic azad university, parsabad moghan, iran). In *3rd International Conference on Foreign Language Teaching and Applied Linguistics*. IBU Publishing, 2013.

[6] C. J. Davis, J. S. Bowers, and A. Memon. Social influence in televised election debates: A potential distortion of democracy. *PloS one*, 6(3):e18154, 2011.

[7] A. De Liddo, S. Buckingham Shum, and M. Klein. Arguing on the web for social innovation: Lightweight tools and analytics for civic engagement. In *Arguing the Web 2.0 Workshop, 8th ISSA Conference on Argumentation*, Amsterdam, 2014.

[8] A. De Liddo, Á. Sándor, and S. Buckingham Shum. Contested collective intelligence: Rationale, technologies, and a human-machine annotation study. *Computer Supported Cooperative Work (CSCW)*, 21(4-5):417–448, 2012.

[9] E. D'heer and C. Courtois. The changing dynamics of television consumption in the multimedia living room. *Convergence: The International Journal of Research into New Media Technologies*, pages 1–16, 2014.

[10] N. A. Diakopoulos and D. A. Shamma. Characterizing debate performance via aggregated twitter sentiment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1195–1198. ACM, 2010.

[11] D. Edge, E. Searle, K. Chiu, J. Zhao, and J. A. Landay. Micromandarin: mobile language learning in context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3169–3178. ACM, 2011.

[12] House of Lords Select Committee on Communications. *Broadcast General Election Debates - HL 171, 2nd Report of Session 2013–14*. The Stationery Office Limited, 2014.

[13] J. B. Lemert. Do televised presidential debates help inform voters? 1993.

[14] P. Lévy and R. Bonomo. *Collective intelligence: Mankind's emerging world in cyberspace*. Perseus Publishing, 1999.

[15] Y.-R. Lin, D. Margolin, B. Keegan, M. Martino, S. Goodman, and D. Lazer. Catching fish in the stream: real time analysis of audience behavior in social media. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*, pages 187–190. ACM, 2013.

[16] T. W. Malone, R. Laubacher, and C. Dellarocas. Harnessing crowds: Mapping the genome of collective intelligence. 2009.

[17] J. A. Martin. Mobile news use and participation in elections: A bridge for the democratic divide? *Mobile Media & Communication*, 3(2):230–249, 2015.

[18] C. Mascaro, A. Black, and S. Goggins. Tweet recall: examining real-time civic discourse on twitter. In *Proceedings of the 17th ACM international conference on Supporting group work*, pages 307–308. ACM, 2012.

[19] M. S. McKinney, J. B. Houston, and J. Hawthorne. Social watching a 2012 republican presidential primary debate. *American Behavioral Scientist*, 58(4):556–573, 2014.

[20] A. Okada and S. Buckingham Shum. Knowledge mapping with compendium in academic research and online education. In *22nd World Conference, International Council of Distance Education*, 2006.

[21] D. Ruths and J. Pfeffer. Social media for large studies of behavior. *Science*, 346(6213):1063–1064, 2014.

[22] A. Selvin, S. Buckingham Shum, M. Sierhuis, J. Conklin, B. Zimmerman, C. Palus, W. Drath, D. Horth, J. Domingue, E. Motta, and G. Li. Compendium: Making meetings into knowledge events. In *Knowledge Technologies*, Austin, Texas, 2001.

[23] A. Smith. Civic engagement in the digital age. *Pew Research Center*, 25, 2013.

[24] M. Xenos, A. Vromen, and B. D. Loader. The great equalizer? patterns of social media use and youth political engagement in three advanced democracies. *Information, Communication & Society*, 17(2):151–167, 2014.